AD-A082 177 CLEMSON UNIV S C DEPT OF MATHEMATICAL SCIENCES F/G 12/1

ON STEIN'S ESTIMATOR.(U)

JUL 76 K ALAM, J S HAWKES N00014-75-C-0451

UNCLASSIFIED N75 NL

END

DATE FILMED

4-80

DTIC

12

6 ON STEIN'S ESTIMATOR.


10 KHURSHEED ALAM*

AND

JAMES S. HAWKES




15 N00014-75-C-0451

11 Jul 76

12 22

9 TECHNICAL REPORT #228

JULY 1976

14 Report N75




*THE AUTHOR'S WORK WAS SUPPORTED BY THE OFFICE OF NAVAL RESEARCH UNDER CONTRACT N00014-75-C-0451.


401183

# ON STEIN'S ESTIMATOR

Khursheed Alam* and James S. Hawkes

Clemson University

## ABSTRACT

Stein's estimator for k normal means is known to dominate the maximum likelihood estimator for $k \geq 3$ if the loss is quadratic. In this paper we have derived certain optimal properties of Stein's estimator for a more general loss function. It is shown that the estimator is minimax in an empirical Bayes sense for the generalized loss function.

Key words and phrases: Maximum Likelihood Estimator; Bayes and Minimax Estimators; Loss Function.

AMS Classification: Primary, 62F10
Secondary, 62,C15

*The author's work was supported by The Office of Naval Research under Contract N00014-75-0451.

1. <u>Introduction</u>. Let $X = (X_1, \ldots, X_k)$ be a vector of k independent observations from k normal populations, where $X_i \overset{d}{\sim} N(\theta_i, 1)$, $i = 1, \ldots, k$. We consider the problem of estimating the vector $\theta = (\theta_1, \ldots, \theta_k)$. Let

$$\delta(X) = (\delta_1(X), \ldots, \delta_k(X))$$

where $\delta_i(X)$ is an estimate of $\theta_i$. Let the loss be squared error, given by

$$L(\theta,\delta) = \frac{1}{k} ||\delta-\theta||^2 = \frac{1}{k} \sum_{i=1}^{k} (\delta_i - \theta_i)^2. \tag{1.1}$$

The maximum likelihood estimator $\delta^0(X) = (X_1, \ldots, X_k)$ has risk equal to

$$R(\theta,\delta^0) = E_\theta\, L(\theta,\delta^0) = 1$$

for each $\theta$. James and Stein (1961) showed that the estimator $\delta^*(X)$, given by

$$\delta_i^*(X) = (1 - \frac{k-2}{S}) X_i \tag{1.2}$$

where $S = ||X||^2$, has risk $R(\theta,\delta^*) < 1$ for all values of $\theta$, if $k \geq 3$. That is, $\delta^*(X)$ dominates $\delta^0(X)$. Therefore, the maximum likelihood estimator is inadmissible. Even though only $X_i$ seems to be relevant to $\theta_i$, it is interesting to note that the estimate $\delta_i^*(X)$ depends on the entire vector of observations $X$.

The surprising result of James and Stein has stimulated considerable research in the last two decades, on the problem

of estimating the mean of a multivariate normal distribution. For reference to some recent work in this area, see Alam (1975) and Berger and Bock (1976). Almost all the known results assume a quadratic loss, chiefly for the reason that nonquadratic loss is difficult to handle mathematically. In this paper we have deviated from the general trend. We consider a more general loss function, given by

$$L(\theta,\delta) = \frac{1}{k}\sum_{i=1}^{k} (\delta_i-\theta_i)^{2m} \tag{1.3}$$

where $m < \frac{k}{2}$ is a positive integer, and obtain the risk of $\delta^*(X)$, using the generalized loss function. In the sequel we shall denote by $R(\theta,\delta)$ the risk of an estimator $\delta(X)$ with respect to the loss (1.3).

The maximum likelihood estimator has a constant risk, given by

$$R(\theta,\delta^0) = \frac{2^m}{\sqrt{\pi}}\,\Gamma(m+\tfrac{1}{2}) \quad . \tag{1.4}$$

On the other hand, $R(\theta,\delta^*)$ is a function of $||\theta||^2$ for $m = 1$. For $m > 1$, let $R^*(\lambda,\delta^*)$ denote the average value of $R(\theta,\delta^*)$ with respect to the distribution of $\theta$, uniform on the sphere

$$||\theta||^2 = \lambda \tag{1.5}$$

with center at the origin and radius equal to $\sqrt{\lambda}$ . In the following section it is shown that

$$R^*(\lambda,\delta^*) < R(\theta,\delta^0) \tag{1.6}$$

for all values of $\lambda$, if $k$ is sufficiently large.

The proof of (1.6) involves very difficult computation. It is perhaps the strongest result we can show for Stein's estimator, since $\delta^*$ is designed for a spherically symmetric loss, whereas (1.3) is spherically non-symmetric. On the other hand, if there exists an estimator which has smaller risk than $\delta^*$ with respect to (1.3), its risk function is likely to be mathematically intractable. Therefore, the inequality (1.6), which is the main result of this paper, is significant, even though we should look for a spherically non-symmetric estimator.

Let $m = 1$. Consider a Bayesian approach. Suppose that the $\theta_i$'s are independently normally distributed with mean zero and variance A. A Bayes estimator of $\theta$ with respect to the given prior is $\hat{\delta}(X)$ where

$$\hat{\delta}_i(X) = (1-B)X_i \ , \quad i=1, \ldots, k \tag{1.7}$$
$$B = 1/(1+A).$$

The Bayes risk is equal to $1-B$.

If B is not known then S is a sufficient statistic for B, and BS is marginally distributed as $\chi_k^2$, chi-squared with k degrees of freedom. Let $\tilde{B}(S)$ be an estimator of B. Substituting $\tilde{B}(S)$ for B in (1.7) we get an empirical Bayes estimator $\tilde{\delta}(X)$, given by

$$\tilde{\delta}_i(X) = (1-\tilde{B}(S))X_i \ , \quad i=1, \ldots, k. \tag{1.8}$$

Let $\rho(B,\tilde{\delta})$ denote the Bayes risk of $\tilde{\delta}$. The difference between $\rho(B,\delta^0)$ and the Bayes risk is equal to B, and

$$\frac{\rho(B,\tilde{\delta}) - \rho(B,\hat{\delta})}{B} = E\left(\frac{\tilde{B}(T)-B}{B}\right)^2 \tag{1.9}$$

where $BT \stackrel{d}{=} \chi_{k+2}^2$.

For $\tilde{B}(S) = \frac{k-2}{S}$ we have $\tilde{\delta}(X) = \delta^{s}(X)$, and the left side of (1.9) is equal to $\frac{2}{k}$. Efron and Morris (1973) have shown that

the minimax value of the right side of (1.9) is equal to $\frac{2}{k}$. Thus Stein's estimator is minimax in an empirical Bayes sense, considering the problem of estimating B with loss function, given by

$$L(B,\tilde{B}) = \left(\frac{\tilde{B}-B}{B}\right)^2 .$$

A similar result on the minimax property of Stein's estimator is obtained for the generalized loss (1.3).

2. Main results. First we compute a Bayes risk of the estimator $\tilde{\delta}(X)$ given by (1.8), using the loss function (1.3). We make the Bayesian assumption $\theta_i \overset{d}{\sim} N(0,A)$ as a convenient mathematical tool. Then marginally, $X_i \overset{d}{\sim} N(0, 1+A)$ and $BS \overset{d}{\sim} \chi_k^2$ .

Let r be a non-negative integer, and let

(2.1) $$d_r = \frac{2^r}{\sqrt{\pi}} \Gamma(r+\tfrac{1}{2})$$

(2.2) $$c_r = \frac{1}{k} E^*((\sum_{i=1}^{k} X_i^{2r})/S^r)$$

$$= \frac{1}{k} (E^*(\sum_{i=1}^{k} X_i^{2r})/(E^* S^r))$$

$$= \frac{2^{-r}\Gamma(\frac{k}{2})}{\Gamma(\frac{k}{2}+r)} d_r$$

where $E^*$ denotes expectation with respect to the marginal distribution of the $X_i$'s. The second line in (2.2) follows

from the fact that the distribution of $(\sum_{i=1}^{k} X_i^{2r})/S^r$ does not depend on A, that S is a complete and sufficient statistic for A, and hence S and $(\sum_{i=1}^{k} X_i^{2r})/S^r$ are independently distributed. Note that $E^* X_i^{2r} = d_r$.

Conditionally, given $X_i$

$$\theta_i | X_i \overset{d}{\sim} N((1-B)X_i, 1 - B) \quad (2.3)$$

Using (2.3) we obtain after simplification, the Bayes risk of $\tilde{\delta}(X)$, given by

$$\begin{aligned} \rho(B,\tilde{\delta}) &= \sum_{r=0}^{m} c_{m-r}\, d_r \binom{2m}{2r} (1-B)^r E^* (S^{m-r} (\tilde{B}(S) - B)^{2m-2r}) \quad (2.4) \\ &= \sum_{r=0}^{m} c_{m-r}\, d_r \binom{2m}{2r} \frac{2^{m-r}\Gamma(\frac{k}{2}+m-n)}{\Gamma(\frac{k}{2})} B^{m-r}(1-B)^r \\ &\qquad E\left\{\frac{\tilde{B}(\chi^2_{k+2m-2r}/B)}{B} - 1\right\}^{2m-2r} \\ &= \sum_{r=0}^{m} a_r\, B^{m-r}(1-B)^r\, E\left\{\frac{\tilde{B}(\chi^2_{k+2m-2r}/B)}{B} - 1\right\}^{2m-2r} \end{aligned}$$

where

$$a_r = c_{m-r}\, d_r \binom{2m}{2r} 2^{m-r}\Gamma(\tfrac{k}{2}+m-r)/\Gamma(\tfrac{k}{2})$$

$$= \frac{2^m}{\pi} \binom{2m}{2r} \Gamma(r+\tfrac{1}{2}) \Gamma(m-r+\tfrac{1}{2})$$

$$= \binom{m}{r} d_m .$$

The Bayes risk of Stein's estimator is obtained by putting $B(S) = \frac{k-2}{S}$ in (2.4). Then

$$(2.5) \quad \rho(B, \delta^*) = \sum_{r=0}^{m} a_r B^{m-r}(1-B)^r E\left(\frac{k-2}{\chi^2_{k+2m-2r}} - 1\right)^{2m-2r}$$

$$= \sum_{r=0}^{m} a_r B^{m-r}(1-B)^r \Phi(-2m+2r, -\frac{k-2}{2} -m+r; -\frac{k-2}{2})$$

$$= \sum_{r=0}^{m} a_r b_r B^{m-r}(1-B)^r$$

where

$$\Phi(a,b;x) = 1 + \frac{a}{b} x + \frac{a(a+1)}{b(b+1)} \frac{x^2}{2!} + \ldots$$

denotes the confluent hypergeometric function, and

$$(2.6) \qquad b_r = \Phi(-2m+2r, -\frac{k-2}{2}- m+r; -\frac{k-2}{2}) .$$

Remark 1: We have that $b_m = 1$ and $b_r \to 0$ as $k \to \infty$ for $r < m$. The confluent hypergeometric function $\Phi(a,b;x)$ satisfies the equation

$$x \frac{d^2y}{dx^2} + (b-x)\frac{dy}{dx} - ay = 0$$

The above equation shows that for $a < 0$, $y > 0$ and $b \leq x < 0$, $y$ is a convex function of $x$ when $\frac{dy}{dx} \geq 0$. Therefore, if $y = 0$ for $x = -\infty$ then $\frac{dy}{dx} \leq 0$ for all $x < 0$. Hence, $b_r$ is decreasing in $k$ for $r < m$, and therefore there is a minimum value of $k = k_m$, say, depending on $m$, such that $b_r \leq 1$ for $r = 0,1,\ldots,m$ and all $k \geq k_m$. Numerical computation shows that $k_m$ is the minimum value of $k$ for which $b_0 \leq 1$. Since

$$\rho(B,\delta^0) = \sum_{r=0}^{m} a_r B^{m-r}(1-B)^r = d_m.$$

we have from (2.5) that

$$\rho(B,\delta^*) \leq \rho(B,\delta^0). \tag{2.7}$$

for $k \geq k_m$ and all values of $B$.

Let $V \overset{d}{=} \chi_k^2(\lambda)$, a non-central chi-squared with $k$ degrees of freedom and non-centrality parameter $\lambda = ||\theta||^2$. The following lemma gives a formula for $R^*(\lambda,\delta^*)$, the average value of $R(\theta,\delta^*)$ on the sphere $||\theta||^2 = \lambda$.

<u>Lemma 2.1</u>.

$$R^*(\lambda,\delta^*) = \frac{\Gamma(\frac{k}{2})}{\Gamma(\frac{k}{2}-m)} \sum_{r=0}^{m} (-1)^r a_r b_r E\left(\Phi(-r,\tfrac{k}{2}-m;\tfrac{V}{2})(\tfrac{V}{2})^{-m}\right) \tag{2.8}$$

$$= \sum_{r=0}^{m} (-1)^r a_r b_r \sum_{t=0}^{r} (-1)^t \binom{r}{t} \Phi(m-t,\tfrac{k}{2};-\tfrac{\lambda}{2}).$$

Proof: Let $U \overset{d}{\sim} \chi_k^2/B$. Apriori, $\theta_i \overset{d}{\sim} N(0,A)$. Therefore $||\theta||^2 \overset{d}{\sim} A\chi_k^2$. The right side of (2.8) is a function of $||\theta||^2$ equal to $g(||\theta||^2)$, say. Taking expectation with respect to the distribution of $||\theta||^2$, we get

(2.9)
$$E\, g(||\theta||^2) = \frac{\Gamma(\frac{k}{2})}{\Gamma(\frac{k}{2}-m)} \sum_{r=0}^{m} (-1)^r a_r b_r \, E((\underline{\Phi}(-r,\tfrac{k}{2}-m;\, \tfrac{U}{2})(\tfrac{U}{2})^{-m})$$
$$= \sum_{r=0}^{m} a_r b_r B^{m-r}(1-B)^r$$
$$= \rho(B,\delta^*) \text{ from (2.5)}\ .$$

Since $\theta_i \overset{d}{\sim} N(0,A)$, the conditional distribution of $\theta$ given $||\theta||^2 = \lambda$ is uniform on the sphere $||\theta||^2 = \lambda$. Therefore

(2.10)
$$\rho(B,\delta^*) = E\, R^*(||\theta||^2,\delta^*)$$

where the expectation on the right side of (2.10) is with respect to the distribution of $||\theta||^2$. Since the distribution of $||\theta||^2$ is complete for B, from (2.9) and (2.10) we have

(2.11)
$$R^*(||\theta||^2,\delta^*) = g(||\theta||^2)$$

for all values of $||\theta||$. From (2.11) we get the first equality in (2.8). The second line in (2.8) is obtained from the first line, using the following formula.

(2.12)
$$E(\tfrac{V}{2})^{-m+t} = e^{-\lambda/2} \sum_{r=0}^{\infty} (\tfrac{\lambda}{2})^r \frac{\Gamma(\frac{k}{2}-m+t+r)}{\Gamma(\frac{k}{2}+r)}$$

$$= e^{-\lambda/2} \frac{\Gamma(\frac{k}{2}-m+t)}{\Gamma(\frac{k}{2})} \bar{\Phi}(\frac{k}{2}-m+t,\ \frac{k}{2};\ \frac{\lambda}{2})$$

$$= \frac{\Gamma(\frac{k}{2}-m+t)}{\Gamma(\frac{k}{2})} \bar{\Phi}(m-t,\ \frac{k}{2};\ -\frac{\lambda}{2}) .$$

The lemma is proved. Q.E.D.

Remark 2: The formula (2.8) can be written as

(2.13)

$$R^*(\lambda,\delta^*) = \sum_{r=0}^{m} (-1)^r a_r b_r \sum_{t=0}^{r} (-1)^t \binom{r}{t} \frac{\Gamma(\frac{k}{2})}{\Gamma(\frac{k}{2}-m+t)} E(\frac{V}{2})^{-m+t}$$

$$= \sum_{t=0}^{m} e_t \, E(\frac{V}{2})^{-m+t}$$

where

(2.14) $$e_t = \sum_{r=t}^{m} (-1)^{r+t} a_r b_r \binom{r}{t} \frac{\Gamma(\frac{k}{2})}{\Gamma(\frac{k}{2}-m+t)} .$$

The values of $e_t$ for $t = m$ and $m - 1$ are computed easily. We have $e_m = d_m$ and

$$e_{m-1} = -\frac{(k-2)^2}{2k} m d_m .$$

Note that the risk of the maximum likelihood estimator is equal to $d_m$ and that for $m = 1$, that is for squared error loss, we have

(2.15) $$R(\theta,\delta^*) = R^*(||\theta||^2,\delta^*)$$

$$= 1 - \frac{(k-2)^2}{k} E\frac{1}{V}$$

$$< 1$$

$$= R(\theta,\delta^0) .$$

We shall now obtain the main result. Let $m > 1$. From the integral representation of the confluent hypergeometric function, given by

$$\Phi(a,b,;x) = \frac{\Gamma(b)}{\Gamma(a)\,\Gamma(b-a)} \int_0^1 e^{xu}\, u^{a-1}(1-u)^{b-a-1}du \qquad b>a>0$$

we find that the value of $\Phi(m-t,\frac{k}{2};-\frac{\lambda}{2})$ lies between 0 and 1, and that it tends to 1 as $k\to\infty$. Since $b_m = 1$ and $b_r \to 0$ as $k\to\infty$ for $r<m$, from (2.8) we have that

$$(2.16) \qquad R^*(\lambda,\delta^*) \to 0 \quad \text{as } k\to\infty .$$

On the other hand, as $\lambda\to\infty$ the summation in the second line on the right side of (2.13) is dominated by the sum of two terms, equal to

$$(2.17) \qquad e_m + e_{m-1}\, E(\tfrac{V}{2})^{-1} = d_m(1-\frac{(k-2)^2}{k}\, E\tfrac{m}{V}).$$

$$< d_m(1-\frac{m(k-2)^2}{k(k+\ )}\ ).$$

Hence

$$(2.18) \qquad R^*(\lambda,\delta^*) \quad d_m$$

for sufficiently large values of $\lambda$.

Next we consider the case in which both k and $\lambda$ are large. Expanding $b_r$, given by (2.6), in negative powers of k we find that $b_m = 1$, $b_{m-1} = \frac{2}{k}$ and $b_r = 0(k^{-2})$ for $r < m-1$. Therefore, for large k, the summation on the right side of (2.8) is dominated by the sum of two terms, equal to D, say where

$$(2.19) \qquad D = \sum_{r=m-1}^{m} (-1)^r a_r b_r \sum_{t=0}^{r} (-1)^t \binom{r}{t} \Phi(m-t,\tfrac{k}{2};-\tfrac{\lambda}{2})$$

$$= e^{-\lambda/2} \sum_{r=m-1}^{m} (-1)^r a_r b_r \sum_{t=0}^{r} (-1)^t \binom{r}{t} \bar{\Phi}\left(\frac{k}{2}-m+t, \frac{k}{2}; \frac{\lambda}{2}\right)$$

$$= E \sum_{r=m-1}^{m} (-1)^r a_r b_r \sum_{t=0}^{r} (-1)^t \binom{r}{t} \frac{\Gamma(\frac{k}{2}-m+t+T)\ \Gamma(\frac{k}{2})}{\Gamma(\frac{k}{2}+T)\ \Gamma(\frac{k}{2}-m+t)}$$

$$= E \sum_{r=m-1}^{m} (-1)^r a_r b_r \sum_{t=0}^{r} (-1)^t \binom{r}{t} \left(\frac{k/2}{T+k/2}\right)^{m-t} \left(1 - \frac{2T}{(T+k/2)k} c_t\right.$$

$$\left. + 0(k^{-2})\right)$$

where T is a random variable distributed according to the Poisson distribution with mean equal to $\lambda/2$ and $c_t = (m-t)(m-t+1)/2$. After simplification we get

$$D = d_m E\left(\frac{T}{T+k/2}\right)^{m-1} \left[\frac{T+2m}{T+k/2} - \frac{m(m+1)k/2}{2(T+k/2)^2}\right] + 0(k^{-2}).$$

$$= d_m E\left(\frac{T}{T+k/2}\right)^{m-1} g_k(T) + 0(k^{-2})$$

where $g_k(T) < 1$ for $k > 4m$. Now

$$E\left(\frac{T}{T+k/2}\right)^{m-1} \leq E\left(\frac{T}{T+k/2}\right)$$

$$\leq 1 - \frac{k/2}{ET+k/2}$$

$$= 1 - \frac{k}{\lambda+k}$$

$$\leq 1 - \frac{1}{1+k} \quad \text{for } \lambda \leq k^2$$

Therefore

$$R^*(\lambda, \delta^*) < d_m$$

for sufficiently large k and $\lambda \leq k^2$. On the other hand, from (2.13) and (2.17) it follows that the above inequality holds also for fficiently large k and $\lambda \geq k^2$. Therefore, we have the following theorem.

<u>Theorem 2.1</u>. Let $\lambda = ||\theta||^2$. Given a positive integer m, there exists a value of $k = k_m^*$, say, depending on m, such that, $R^*(\lambda, \delta^*) < d_m = R(\theta, \delta^0)$ for all $\theta$ and $k \geq k_m^*$.

Next we show that Stein's estimator is minimax in an empirical Bayes sense. From (2.4) we have

$$(2.20) \qquad \rho(B, \tilde{\delta}) = d_m \sum_{r=0}^{m} \binom{m}{r} B^r (1-B)^{m-r} E\left(\frac{\tilde{B}(W_r)}{B} - 1\right)^{2r}$$

where $W_r \overset{d}{\sim} \chi^2_{k+2r}/B$. In deriving the above result we have assumed that $\theta_i \overset{d}{\sim} N(0,A)$. We can stop short of full Bayesian-hood by assuming that A or equivalently B, is unknown and must be estimated. That is, we become empirical Bayesian, considering $\tilde{B}(W_r)$ as an estimate of B.

Consider a beta prior distribution on B with density function

$$g(B) = B^{\xi-1}(1-B)^{\eta-1} \frac{\Gamma(\xi+\eta)}{\Gamma(\xi)\Gamma(\eta)} \quad , \qquad \xi>0, \ \eta>0.$$

Let

$$\Omega(B, \tilde{\delta}) = (\rho(B, \tilde{\delta}) - (1-B)^m d_m)/B \quad .$$

From (2.20) we have

(2.21)

$$\int_0^1 \Omega(B,\tilde{\delta})g(B)dB \geq \sum_{r=1}^{m} d_m \int_0^1 \binom{m}{r} B^{r-1}(1-B)^{m-r}\left(\frac{\hat{B}(W_r)}{B} - 1\right)^{2r} g(B)dB$$

where

$$\hat{B}(W_r) = \frac{\int_0^1 B^{\frac{k}{2}+\xi-1}(1-B)^{m-r+\eta-1}e^{-BW_r/2}dB}{\int_0^1 B^{\frac{k}{2}+\xi-2}(1-B)^{m-r+\eta-1}e^{-BW_r/2}dB}$$

$$= \frac{(\frac{k}{2}+\xi-1)\overline{\Phi}(\frac{k}{2}+\xi,\frac{k}{2}+m-r+\xi+\eta;\ -W_r/2)}{(\frac{k}{2}+m-r+\xi+\eta-1)\overline{\Phi}(\frac{k}{2}+\xi-1,\frac{k}{2}+m-r+\xi+\eta-1;\ -W_r/2)}.$$

From the asymptotic property of the confluent hypergeometric function (see e.g. Erdelyi (1953) § 6.13.1) it is seen that

(2.22) $$\nu(W_r) = \frac{W_r \hat{B}(W_r)}{k-2} \to \frac{k+2\xi-2}{k-2} \quad \text{as } W_r \to \infty.$$

Let $\xi \to 0$. Then B converges in distribution to zero and $W_r$ converges in distribution to $\infty$. Therefore, $\nu(W_r)$ converges in distribution to 1, and from (2.21) we have

(2.23)

$$\operatorname*{Sup}_B \Omega(B,\tilde{\delta}) \geq \lim_{\xi\to 0} \int_0^1 \Omega(B,\tilde{\delta})g(B)dB$$

$$\geq \lim_{\xi\to 0} d_m \int_0^1 \left(\sum_{r=1}^{m} \binom{m}{r} B^{r-1}(1-B)^{m-r} E\left(\frac{k-2}{BW_r}\nu(W_r)-1\right)^{2r} g(B)dB\right.$$

$$= \operatorname*{Lim}_{\xi \to 0} d_m \int_0^1 \left(\sum_{r=1}^{m} \binom{m}{r} B^{r-1} (1-B)^{m-r} E\left(\frac{k-2}{BW_r} - 1\right)^{2r} g(B)\,dB\right.$$

$$= \operatorname*{Lim}_{\xi \to 0} \int_0^1 \Omega(B, \delta^*) g(B)\,dB$$

$$= a_{m-1}\, b_{m-1}$$

$$= \frac{2m}{k}\, d_m$$

$$= \operatorname*{Sup}_{B} \Omega(B, \delta^*) \text{ for } k \geq k_m^{**}, \text{ say.}$$

The last equality in (2.23) follows from the fact that $b_r = 0(k^{-2})$ for $r < m - 1$. Thus we have obtained the following result from the empirical Bayes approach, considered above.

<u>Theorem 2.2</u>. The minimax value of the Bayes risk $\rho(B, \tilde{\delta})$ is equal to $\frac{2m}{k} d_m = \operatorname*{Sup}_{B} \Omega(B, \delta^*)$ for $k > k_m^{**}$.

<u>Remark 3</u>: Clearly, the results given above are applicable to the case in which the loss is a polynomial function of the squared error with non-negative coefficients. It might be possible to extend the given results to the case in which the loss is a monotone increasing convex function of the squared error.

Table I below gives values of $k_m^*$ and $k_m^{**}$ for $m = 1(1)10$. For the given values of $m$ it is found that $k_m^* = k_m$ where $k_m$ is the smallest value of $k$ for which (2.7) holds. Table II below gives values of $R^*(\lambda, \delta^*)/d_m$ for $m = 1(1)10$ and $\lambda$ and $k = k_m^*$.

Note that $R(\theta,\delta^0) = d_m$. The asymptotic formula for large negative values of the argument of the confluent hypergeometric function, given by

$$\Phi(a,b;-x) = \frac{\Gamma(b)}{\Gamma(b-a)} x^{-a}(1+O(x^{-1}))$$

was used in the computation of $R^*(\lambda,\delta^*)$ for $\lambda = 100, 1000$.

TABLE I - Values of $k_m^*$ and $k_m^{**}$

| m | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $k_m^*$ | 3 | 5 | 7 | 10 | 12 | 15 | 17 | 20 | 22 | 25 |
| $k_m^{**}$ | 3 | 5 | 8 | 10 | 13 | 15 | 18 | 21 | 24 | 27 |

TABLE II - Values of $R^{*}(\lambda,\delta^{*})/d_m$ for $k = k_m^{*}$

| $\lambda$ \ m | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | .6666 | .5714 | .8581 | .5301 | .8640 | .5758 | .9178 | .6247 | .9779 | .6724 |
| 1 | .7584 | .5294 | .6444 | .4004 | .6176 | .4202 | .6532 | .4522 | .6959 | .4851 |
| 2 | .8206 | .5239 | .5137 | .3140 | .4513 | .3119 | .4709 | .3311 | .5000 | .3531 |
| 3 | .8634 | .5379 | .4377 | .2573 | .3385 | .2360 | .3444 | .2455 | .3631 | .2594 |
| 4 | .8933 | .5615 | .3976 | .2212 | .2620 | .1826 | .2560 | .1845 | .2666 | .1925 |
| 5 | .9145 | .5889 | .3808 | .1997 | .2101 | .1447 | .1937 | .1407 | .1981 | .1444 |
| 6 | .9299 | .6170 | .3790 | .1886 | .1753 | .1179 | .1494 | .1089 | .1491 | .1095 |
| 7 | .9412 | .6440 | .3866 | .1849 | .1525 | .0989 | .1178 | .0858 | .1138 | .0841 |
| 10 | .9614 | .7132 | .4357 | .2008 | .1264 | .0703 | .0669 | .0468 | .0554 | .0410 |
| 15 | .9757 | .7899 | .5302 | .2640 | .1454 | .0672 | .0442 | .0257 | .0236 | .0165 |
| 20 | .9823 | .8361 | .6066 | .3346 | .1899 | .0863 | .0477 | .0231 | .0161 | .0097 |
| 25 | .9860 | .8661 | .6642 | .3995 | .2405 | .1152 | .0618 | .0278 | .0160 | .0082 |
| 100 | .9966 | .9642 | .8958 | .7640 | .6397 | .4744 | .3515 | .2207 | .1410 | .0720 |
| 1000 | .9996 | .9964 | .9893 | .9746 | .9589 | .9340 | .9105 | .8769 | .8468 | .8061 |

## REFERENCES


[1] Alam, K. (1975). Minimax and admissible minmax estimators of the mean of a multivariate normal distribution for unknown covariance matrix. Journal of Multivariate Analysis. (5) 83-95.

[2] Berger, James and Bock, M. E. (1976). Combining independent normal mean estimation problems with unknown variances. Ann. Statist. (4) 642-648.

[3] Efron, B. and Morris, C. (1973). Stein's estimation rule and its competitors - an empirical Bayes approach. Journal of the American Statistical Association, (68) 117-130.

[4] Erdelyi, A. (1953). Higher Transcendental Functions. In Bateman Manuscript Project, Volume I.

[5] James, W. and Stein, C. (1961). Estimation with quadratic loss. Proceedings of the Fourth Berkeley Symposium, University of California Press (1) 361-79.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

# REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

| 1. REPORT NUMBER | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
|---|---|---|
| N-75 | | |

| | |
|---|---|
| 4. TITLE (and Subtitle)<br>On Stein's Estimator | 5. TYPE OF REPORT & PERIOD COVERED<br>TR #228 |
| | 6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s)<br>Khursheed Alam<br>James S. Hawkes | 8. CONTRACT OR GRANT NUMBER(s)<br>N00014-75-C-0451 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Clemson University<br>Dept. of Mathematical Sciences<br>Clemson, South Carolina 29631 | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS<br>NR 042-271 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>Office of Naval Research<br>Code 436<br>Arlington, Va. 22217 | 12. REPORT DATE<br>7/76 |
| | 13. NUMBER OF PAGES<br>18 |
| 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) | 15. SECURITY CLASS. (of this report)<br>Unclassified |
| | 15a. DECLASSIFICATION DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Maximum Likelihood Estimator; Bayes and Minimax Rules; Loss Function.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Stein's estimator for k normal means is known to dominate the maximum likelihood estimator for $k \geq 3$ and quadratic loss function. In this paper are given certain optimal properties of Stein's estimator for a generalized loss function.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)